Human Motion Diffusion Model

Abstract

Natural and expressive human motion generation is the holy grail of computer animation. It is a challenging task, due to the diversity of possible motion, human perceptual sensitivity to it, and the difficulty of accurately describing it. Therefore, current generative solutions are either low-quality or limited in expressiveness. Diffusion models, which have already shown remarkable generative capabilities in other domains, are promising candidates for human motion due to their many-to-many nature, but they tend to be resource hungry and hard to control. In this paper, we introduce Motion Diffusion Model (MDM), a carefully adapted classifier-free diffusion-based generative model for the human motion domain. MDM is transformer-based, combining insights from motion generation literature. A notable design-choice is the prediction of the sample, rather than the noise, in each diffusion step. This facilitates the use of established geometric losses on the locations and velocities of the motion, such as the foot contact loss. As we demonstrate, MDM is a generic approach, enabling different modes of conditioning, and different generation tasks. We show that our model is trained with lightweight resources and yet achieves state-of-the-art results on leading benchmarks for text-to-motion and action-to-motion. https://guytevet.github.io/mdm-page/ .